Praxis für Humangenetik Tübingen
Dr. med. Dr. rer. nat. Saskia Biskup
Dr. med. Konstanze Hörtnagel
Fachärzte für Humangenetik

Variant Classification Criteria

General information

Variant pathogenicity classification performed at Praxis for Humangenetik Tübingen is carried out using, but is not limited to, the following sources of information.

1. HGVS sequence variant nomenclature is utilised throughout the process, using standardised gene codes (HGNC).
2. Described disease and gene information; phenotype catalogues in use include OMIM (www.omim.org), Orphanet (www.orpha.net), and GeneReviews (http://www.ncbi.nlm.nih.gov/books/NBK1116/) to give indications of mechanisms of disease, inheritance patterns, phenotypic spectra, expressivity and penetrance.
3. Population frequency; variant population frequency databases in use include dbSNP (www.ncbi.nlm.nih.gov/projects/SNP/), Exome Variant Server (evs.gs.washington.edu/), ExAC (exac.broadinstitute.org/), and an in-house database. Database of Genomic Variants (DGV; http://dgv.tcag.ca/dgv/app/home) is used for array comparative genomic hybridisation data.
4. Pathogenic variant databases, including HGMD (http://www.hgmd.cf.ac.uk/), LOVD (http://databases.lovd.nl/shared/genes), DECIPHER (https://decipher.sager.ac.uk), or ClinVar (http:ncbi.nlm.nih.gov/clinvar), or gene specific databases (e.g www.cftr2.org, http://www.gzneurosci.com/scn1adatabase/), taking into account known genotype-phenotype correlations.
5. Relevant literature.
6. *In silico* prediction algorithms, including; Mutation Taster, fathmm, Mutation Assessor, SIFT, fathmm-MKL coding, LRT, and PROVEAN for missense variants. Splice region analysis (-3 to -8 of the 5' splice site and -12 to +2 of the 3' splice site) is performed using using two algorithms, ADA and RF (Jian et al., 2014, PMID 25416802), and all synonymous variants are analysed using NNSplice (http://www.fruitfly.org/seq_tools/splice.html) and NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/).
7. Genomic sequence evolutionary conservation utilising the 100 vertebrates basewise conservation by PhyloP score (https://genome.ucsc.edu/cgi-bin/hgTracks).
8. Protein functional domains or published molecular structure; UniProt (http://www.uniprot.org/), NextProt (http://www.nextprot.org/).
9. Segregation with disease within related individuals.
10. Allele state relative to other identified pathogenic variants; e.g cis/trans configuration.

Classification

Variant classification is performed with close adherence to ACMG guidelines (Richards et al., 2015; PMID: 25741868). Each variant is analysed according to multiple criteria, and categorised according to multiple supporting sources. Variants are classified into five categories; benign, likely benign, unclear, likely pathogenic, and pathogenic.

Benign

Benign variants are classified based upon a very unlikely link to disease. These variants have no supporting evidence for a link to disease, these include:

- Population frequency > 1% for recessive or X-linked recessive variants, or greater than > 0.1% for dominant and X-linked dominant variants for rare disorders, unless otherwise explicitly indicated in relevant literature.
- Allele frequency is higher than would be expected for the specific disorder.
- Described benign within literature.

Likely Benign

The role of likely benign variants in disease pathogenesis cannot be safely ruled out, however they are unlikely to be causative. These include;

- Variants identified in trans state in conjunction with a likely, or known pathogenic variant within a gene known to have a fully penetrant autosomal dominant inheritance pattern.
- In-frame or splice region indels within a tandem repeat region, without an explicit link to disease pathogenesis.
- Variants identified within gene regions demonstrated to not have high sequence constraints.
- Synonymous or splice region excluding canonical splice variants (> +2 and < -2), without a splice changing prediction and low sequence conservation.
- Variant identified in a case with a clear alternative cause of disease.
- Variant or variants identified in the same allele state in a healthy individual.

Unclear

These variants have no direct supporting information regarding their role in disease pathogenesis, yet have the potential to be causative. This is the default variant annotation in the absence of further evidence, this is inclusive where variants meet no criteria, and in cases where conflicting information has been identified. These include;

- Variants within gene linked to relevant disease, without familial segregation information from multiple affected and unaffected individuals, including variants predicted to be pathogenic using in silico prediction tools.

- Variants at the same position as a known pathogenic variant, but without a “pathogenic” prediction, or at a position of low sequence conservation.
- Variants located within a gene region without a clear or well established functional link to disease.
- Lack of sufficient evidence within the original publication, e.g. identification of a single variant within a large cohort with no segregation information or functional analysis.
- Lack of detailed patient clinical information linking the disease with described clinical information.
- In-frame insertion or deletion.

Likely Pathogenic

Likely pathogenic variants show a very likely association with the disease pathogenesis, including expected loss of function variants, however there are as yet no direct links to disease pathogenesis, including;

- Frameshift and nonsense variants which are likely to be loss-of-function (not at extreme 3’ end of gene), canonical splice site variants, initiation codon change variants, or CNVs each with disease-relevant inheritance patterns, which have not yet been described in the literature, only in cases in which a loss-of-function variant is a known pathomechanism.
- In-frame variant covering a region where an in-frame variant has been described.
- Missense variants within hot-spot regions, with in-silico prediction tools giving a ‘pathogenic’ prediction, and within a region with high sequence conservation; or at the same amino acid position as a known pathogenic variant, within gene consistent with disease of patient
- De novo variants within a gene consistent with disease of patient, with a ‘pathogenic’ prediction, and high sequence conservation.
- Missense variants resulting in the same amino acid substitution as a known described variant.
- Functional studies, in vivo or in vitro, which indicate a clearly damaging effect of the variant within the gene product and an existing link to known disease.
- Variant in trans configuration with a known pathogenic variant within a gene with a clear link to autosomal recessive disease indicated in the patient, with a pathogenic prediction and high sequence conservation.

Pathogenic

Pathogenic variants are known to be causative for the described genetic disorder, or are very likely to be causative based upon multiple sources of complementing information, including;

- Known described pathogenic variant in the described disorder, with an inheritance pattern consistent with the described disease.